CompSci 295, Causal Inference

Rina Dechter, UCI

Lecture 5b: Linear Structural Causal Models

Slides: Daniel Kumor, Elias Bareinboim

Lecture Outline

1. Introduction to Linear Structural Causal Models
2. Examples of when regression can and cannot be used to find causal effects.
3. Modern algorithmic approaches to identification in linear SCM

Linear Structural Causal Models

Linear SCM are defined as a system of linear equations representing ground-truth:

$$Y := \sum_i \lambda_{x_i y} X_i + \mathcal{E}_y$$

1. All correlations between $\mathcal{E}$ are explicitly specified.
2. X_i are the direct causes of Y, and $\lambda_{x_i y}$ is the change in Y per X_i.
3. WLOG assume normalized data ($\mathbf{E}[X] = 0$ and $\mathbf{E}[XX] = 1$) to simplify math
4. Assume $\mathcal{E}_y \sim \mathcal{N}$, meaning that the distribution is fully specified by covariance matrix Σ (σ_{ij}).

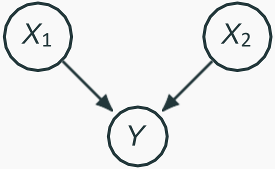
Non-Parametric to Linear

The only substantive change we are making is that the function f becomes linear:

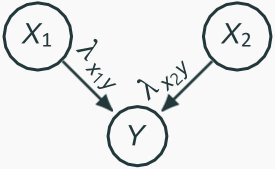
$$V_i \leftarrow f_i(pa_i, U_i) \quad \Rightarrow \quad V_i \leftarrow \sum_{j|V_j \in pa_i} \lambda_{ji} V_j + \mathcal{E}_i$$

1. λ_{ji} is called the "Structural Coefficient".
2. Instead of using U_i, we rename it to $\mathcal{E}_i$ by convention.
3. If we know all λ_{ji}, we can find the causal effect of V_j on V_i.

Example



$\Longrightarrow$
becomes



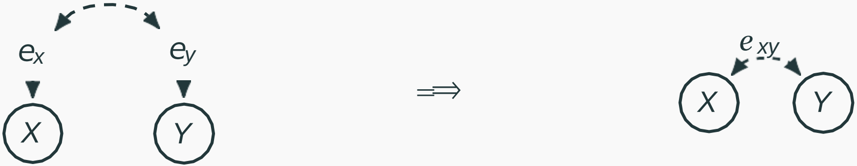
$$
\begin{aligned}
X_1 &= f_{x_1}(U_{x_1}) \\
X_2 &= f_{x_2}(U_{x_2}) \\
Y &= f_y(X_1, X_2, U_y)
\end{aligned}
$$

$$
\begin{aligned}
X_1 &= E_{x_1} \\
X_2 &= E_{x_2} \\
Y &= \lambda_{x_1y}X_1 + \lambda_{x_2y}X_2 + E_y
\end{aligned}
$$

We can draw the structural coefficients directly on the graph, which then fully specifies the model.

Latent Confounding

The covariance between e_i and e_j is represented by e_{ij}, and is used as the value of a bidirected edge:

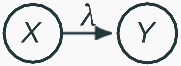


$$e_{xy} \equiv \mathbb{E}[e_x e_y]$$

e_{xy} is unobserved, since it is covariance of latent variables. It is mathematically useful, however, so we draw it on the graph just like structural coefficients.

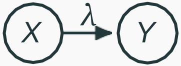
This is different from graph of non-parametric SCM, where a bidirected edge represents an explicit latent variable.

Linear SCM: Interventions



$$\mathbf{E}[Y | do(X = x)] = ?$$

Linear SCM: Interventions



$$
\begin{aligned}
\mathbf{E}[Y \mid do(X = x)] &= \mathbf{E}[\lambda x + e_y] \\
&= \lambda x + \mathbf{E}[e_y] \\
&= \lambda x
\end{aligned}
$$

Identification In Linear SCM: The Problem Statement

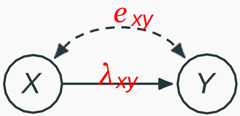
- **Graph**: We are assuming that you have a hypothesized causal graph structure. In other words, you think you know what causes what, and which variables have an unknown common cause.
- **Observational Data**: You have a set of datapoints with measurements of all of the observable variables.
- **Goal: Structural Coefficients** You do NOT have knowledge of the underlying structural coefficients. These represent the actual causal effects that we want to find.

(x_1, y_1)

(x_2, y_2)

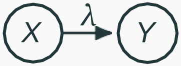
...

(x_n, y_n)



Connecting Observed with Unobserved

Remember that we assumed $e \sim N$, meaning that the distribution is fully specified by covariance matrix Σ (σ_{ij}).



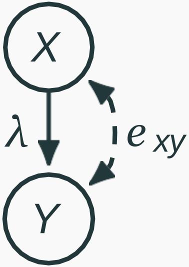
Remember, we normailize
The mean to 0 and variance to 1

$$
\begin{aligned}
\sigma_{xy} &= \mathbf{E}[XY] \\
&= \mathbf{E}[X(\lambda X + e_y)] \\
&= \mathbf{E}[\lambda XX + X e_y] \\
&= \lambda \mathbf{E}[XX] + \mathbf{E}[X e_y] \\
&= \lambda 1 + 0 \\
&= \lambda
\end{aligned}
$$

Connecting Observed with Unobserved

Solve for σ_{xy} in terms of the structural coefficients λ and e $_{xy}$

.

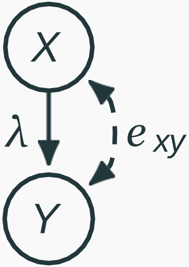


$\sigma_{xy} = ?$

Connecting Observed with Unobserved

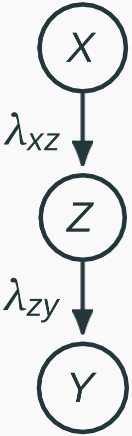
Solve for σ_{xy} in terms of the structural coefficients λ and e $_{xy}$

.



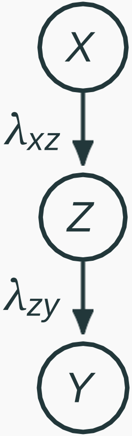
$$
\begin{aligned}
\sigma_{xy} &= \mathbf{E}[XY] \\
&= \mathbf{E}[X(\lambda X + e_y)] \\
&= \mathbf{E}[\lambda XX + X e_y] \\
&= \lambda \mathbf{E}[XX] + \mathbf{E}[X e_y] \\
&= \lambda 1 + \mathbf{E}[X e_y] \\
&= \lambda 1 + \mathbf{E}[e_x e_y] \\
&= \lambda + \; e_{xy}
\end{aligned}
$$

A Curious Property



$\sigma_{xy} = ?$

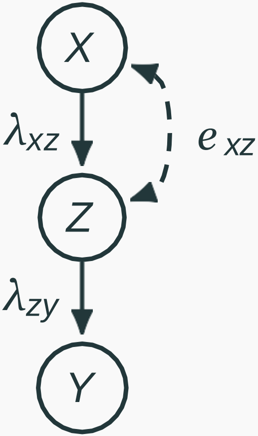
A Curious Property



$$
\begin{aligned}
\sigma_{xy} &= \mathbf{E}[XY] \\
&= \mathbf{E}[X(\lambda_{zy}Z + e_y)] \\
&= \mathbf{E}[\lambda_{zy}XZ + Xe_y] \\
&= \lambda_{zy}\mathbf{E}[XZ] + \mathbf{E}[Xe_y] \\
&= \lambda_{zy}\mathbf{E}[XZ] \\
&= \lambda_{zy}\mathbf{E}[X(\lambda_{xz}X + e_z)] \\
&= \lambda_{zy}\lambda_{xz}\mathbf{E}[XX] + \lambda_{zy}\mathbf{E}[Xe_z] \\
&= \lambda_{zy}\lambda_{xz}
\end{aligned}
$$

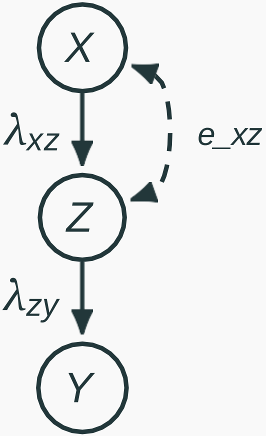
We replace X with e_x

A Curious Property



$\sigma_{xy} = ?$

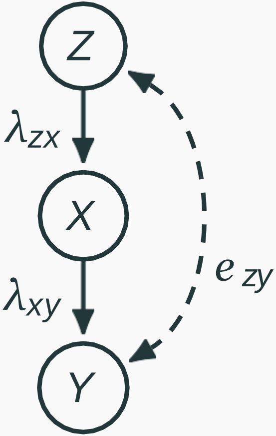
A Curious Property



$$
\begin{aligned}
\sigma_{xy} &= \mathbf{E}[XY] \\
&= \mathbf{E}[X(\lambda_{zy}Z + e_y)] \\
&= \mathbf{E}[\lambda_{zy}XZ + Xe_y] \\
&= \lambda_{zy}\mathbf{E}[XZ] + \mathbf{E}[Xe_y] \\
&= \lambda_{zy}\mathbf{E}[XZ] \\
&= \lambda_{zy}\mathbf{E}[X(\lambda_{xz}X + e_z)] \\
&= \lambda_{zy}\lambda_{xz}\mathbf{E}[XX] + \lambda_{zy}\mathbf{E}[Xe_z] \\
&= \lambda_{zy}\lambda_{xz} + \lambda_{zy}\ e_{xz}
\end{aligned}
$$

Paths & Covariances

There seems to be a relationship between covariances and paths in the graph.



$$\sigma_{xy} = \mathbf{E}[XY] = \mathbf{E}[X(\lambda_{xy}X + e_y)]$$

$$= \lambda_{xy}\mathbf{E}[XX] + \mathbf{E}[Xe_y]$$

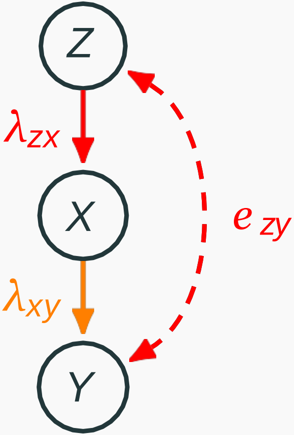
$$= \lambda_{xy} + \mathbf{E}[(\lambda_{zx}Z + e_x)e_y]$$

$$= \lambda_{xy} + \lambda_{zx}\mathbf{E}[e_z e_y] + \mathbf{E}[e\_x, e\_y]$$

$$= \lambda_{xy} + \lambda_{zx}\ e_{zy}$$

Paths & Covariances

There seems to be a relationship between covariances and paths in the graph.



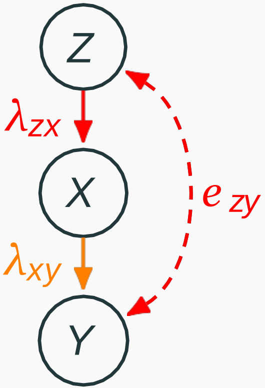
$$\sigma_{xy} = \lambda_{xy} + \lambda_{zx}\ e_{zy}$$

The resulting terms correspond to paths between X and Y in the causal graph

Treks & Wright's Rule

The covariance between variables X and Y is the sum of paths between them in the causal graph, i.e. any non-self-intersecting path without colliding arrowheads ($\rightarrow\leftarrow$):

$x \leftarrow \ldots \leftrightarrow \ldots \rightarrow y \qquad x \leftarrow \ldots \leftarrow w \rightarrow \ldots \rightarrow y \qquad x \leftarrow \ldots \leftarrow y \qquad x \rightarrow \ldots \rightarrow y$



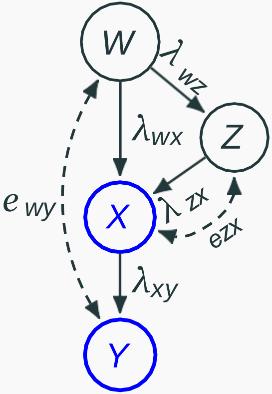
$$\sigma_{xy} = (X \xrightarrow{\lambda_{xy}} Y) + (X \xleftarrow{\lambda_{zx}} Z \xleftrightarrow{e\_zy} Y)$$

$$\sigma_{xy} = \lambda_{xy} + \lambda_{zx}\ e_{zy}$$

Reading Covariances off the Graph

The covariance between variables X and Y is the sum of open paths between them in the causal graph, so paths with no colliding arrowheads ($\rightarrow\leftarrow$):

$$x \leftarrow \ldots \leftrightarrow \ldots \rightarrow y \qquad x \leftarrow \ldots \leftarrow w \rightarrow \ldots \rightarrow y \qquad x \leftarrow \ldots \leftarrow y \qquad x \rightarrow \ldots \rightarrow y$$

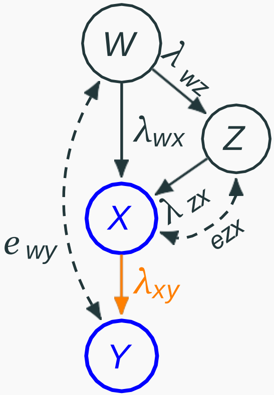


$$\begin{aligned} \sigma_{xy} = & \quad \lambda_{xy} \\ & + \lambda_{wx}\, e_{wy} \\ & + \lambda_{zx}\lambda_{wz}\, e_{wy} \end{aligned}$$

Reading Covariances off the Graph

The covariance between variables X and Y is the sum of open paths between them in the causal graph, so paths with no colliding arrowheads ($\rightarrow\leftarrow$):

$$x \leftarrow \dots \leftrightarrow \dots \rightarrow y \qquad x \leftarrow \dots \leftarrow w \rightarrow \dots \rightarrow y \qquad x \leftarrow \dots \leftarrow y \qquad x \rightarrow \dots \rightarrow y$$

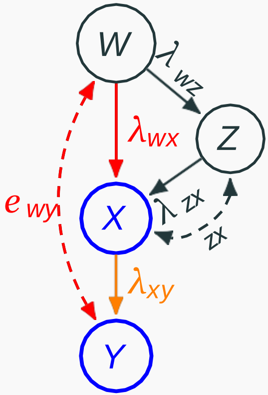


$$\sigma_{xy} = \lambda_{xy} + \lambda_{wx}\, e_{wy} + \lambda_{zx}\lambda_{wz}\, e_{wy}$$

Reading Covariances off the Graph

The covariance between variables X and Y is the sum of open paths between them in the causal graph, so paths with no colliding arrowheads (→←):

$$x \leftarrow \dots \leftrightarrow \dots \rightarrow y \qquad x \leftarrow \dots \leftarrow w \rightarrow \dots \rightarrow y \qquad x \leftarrow \dots \leftarrow y \qquad x \rightarrow \dots \rightarrow y$$

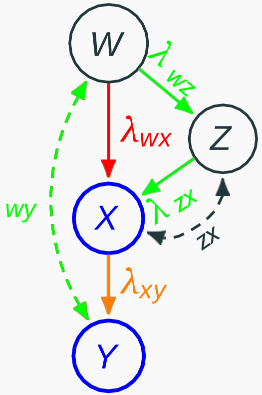


$$\sigma_{xy} = \lambda_{xy} + \lambda_{wx\ wy} + \lambda_{zx}\lambda_{wzEwy}$$

Reading Covariances off the Graph

The covariance between variables X and Y is the sum of open paths between them in the causal graph, so paths with no colliding arrowheads ($\rightarrow\leftarrow$):

$$x \leftarrow \dots \leftrightarrow \dots \rightarrow y \qquad x \leftarrow \dots \leftarrow w \rightarrow \dots \rightarrow y \qquad x \leftarrow \dots \leftarrow y \qquad x \rightarrow \dots \rightarrow y$$



$$\sigma_{xy} = \lambda_{xy} + \lambda_{wx}\ {wy} + \lambda_{zx}\lambda_{wz}\ e\ {wy}$$

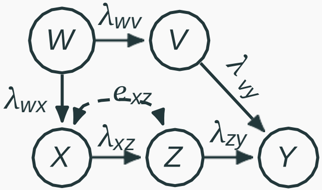
Wright's Rules (1921)

Wright's Rules [9]

$$\sigma_{xy} = \text{Sum of products of path coefficients}$$
$$\text{along all open paths between } X \text{ and } Y$$

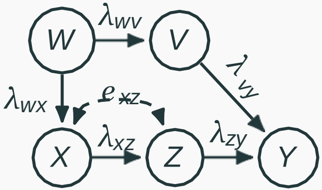
- σ_{xy} is only 0 when X and Y are d-separated.
- If there is an edge $X \xrightarrow{a} Y$ in the model, then
  $\sigma_{xy} = \alpha +$ other paths between x and y.
  Thus $\sigma_{xy} = \alpha$ if X and Y are d-separated in G_α (graph where edge α is removed)
- Wright's rules are defined for acyclic models

One More Example



$$\sigma_{xy} = ?$$

One More Example



$$\sigma_{xy} = (\lambda_{xz} + e_{xz})\lambda_{zy} + \lambda_{wx}\lambda_{wv}\lambda_{vy}$$

Linear Regression

Example: The Medical Researcher

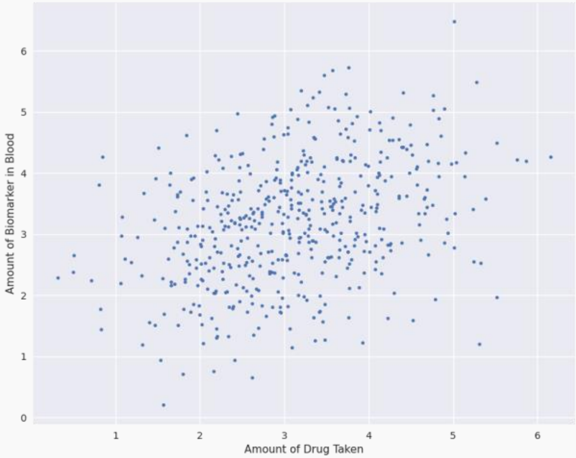
Suppose we are a medical researchers who are trying to determine if a new drug is helpful for curing a disease.

Example: The Medical Researcher

Suppose we are a medical researchers who are trying to determine if a new drug is helpful for curing a disease.

Our job is to make a treatment recommendation, which will be followed by doctors around the country.

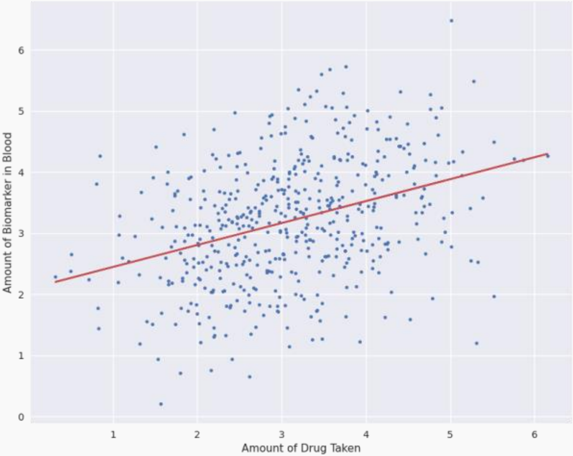
Step 1: Gather a Dataset



Start by gathering a dataset of patients who have taken the drug, including:

1. How much of the drug they took
2. The amount of a biomarker (antibodies) in their blood.

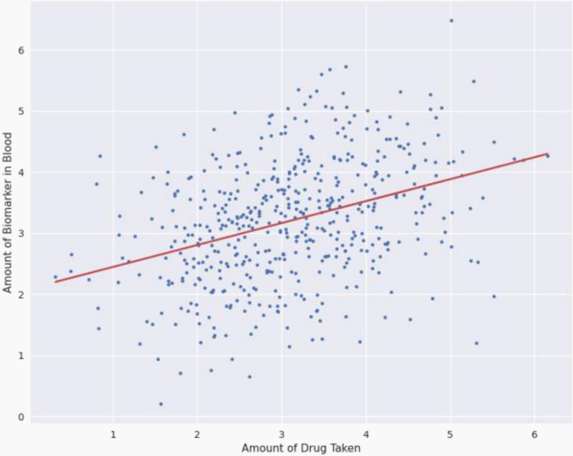
Step 2: Perform a Regression



Perform a regression $Y = \beta X + e$ on the data, with X as amount of drug taken, and Y the amount of biomarker, giving:

$$\beta = 0.375$$

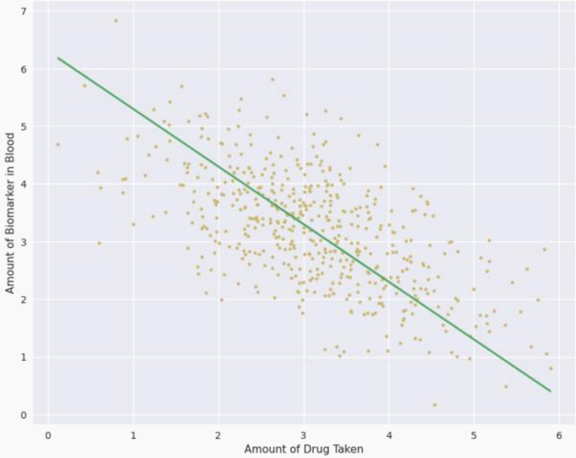
Step 2: Perform a Regression



Perform a regression $Y = \beta X + e$ on the data, with X as amount of drug taken, and Y the amount of biomarker, giving: $\beta = 0.375$

The drug seems to be beneficial, so you authorize its use.

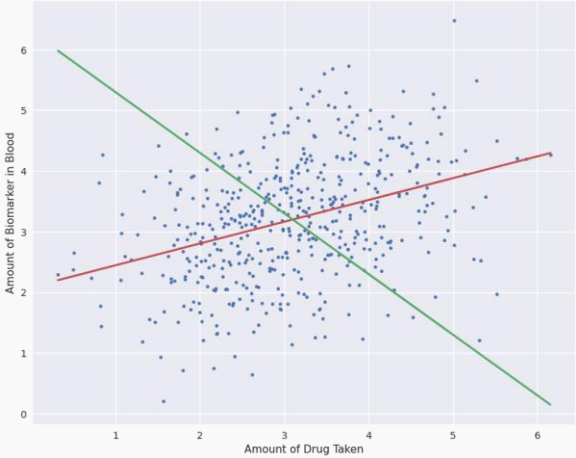
Step 3: The Drug is Given to Everyone



When the drug is given to everyone in the population, the result is a clear negative association, with slope -1.

This drug actually hurts people!

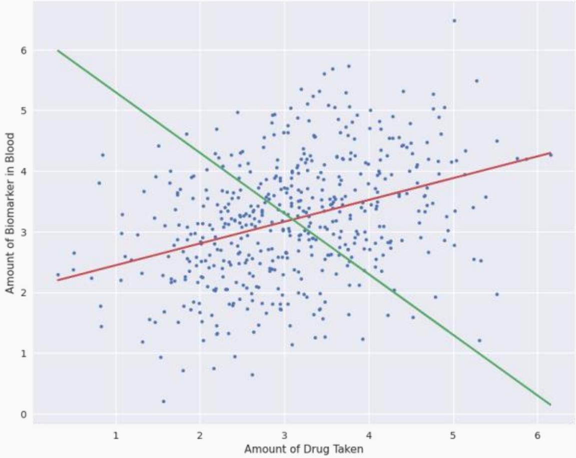
What's Happening Here?



Why was this negative effect not visible in the original dataset?

- Maybe we didn't gather enough data?

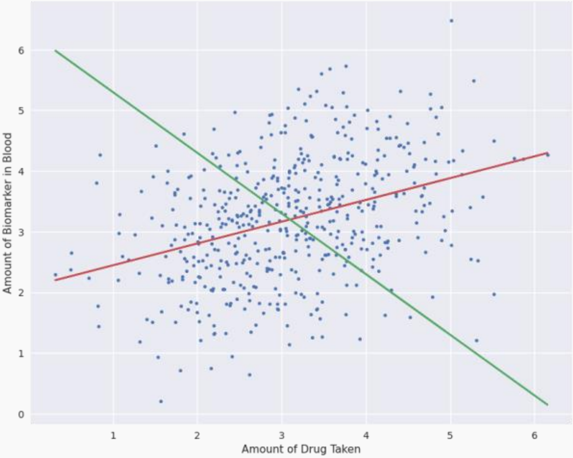
What’s Happening Here?



Why was this negative effect not visible in the original dataset?

- ~~Maybe we didn’t gather enough data?~~
- Why did the original regression “fail" here? (red line)

What's Happening Here?

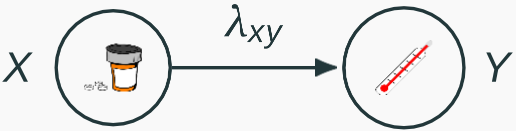


Why was this negative effect not visible in the original dataset?

- ~~Maybe we didn't gather enough data?~~
- Why did the original regression "fail" here? (red line)
- Is there a way to get the true causal effect? (green line)

Key Assumption: Lack of Confounding

The following world model is implicitly assumed when attributing causal meaning to the regression coefficient:

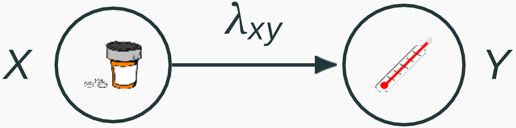


$$X := e_x$$
$$Y := \lambda_{xy} X + e_y$$

e_x, e_y independent

Key Assumption: Lack of Confounding

The following world model is implicitly assumed when attributing causal meaning to the regression coefficient:



$$X := e_x$$

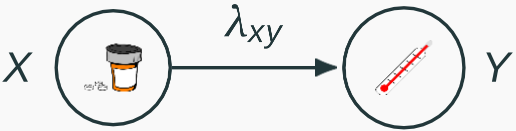
$$Y := \lambda_{xy} X + e_y$$

e_x, e_y independent

Regression $Y = \beta X + e$ gives correct $\beta = \lambda_{xy}$.

Key Assumption: Lack of Confounding

The following world model is implicitly assumed when attributing causal meaning to the regression coefficient:



$$X := e_x$$

$$Y := \lambda_{xy} X + e_y$$

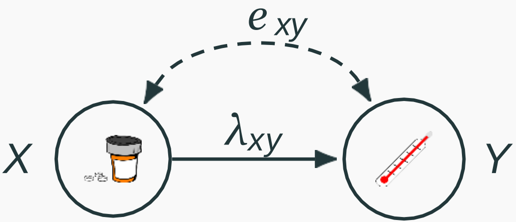
e_x, e_y independent

The covariance gives the same answer:

$$\sigma_{xy} = \mathrm{E}[XY] = \mathrm{E}[X(\lambda_{xy} X + e_y)] = \lambda_{xy} \overbrace{\mathrm{E}[XX]}^{1} + \overbrace{\mathrm{E}[X e_y]}^{0}$$

The Ground-Truth Model

If one is unable to ascertain the assumption of no confounding between X and Y, this is the corresponding graphical model:



$$X := e_x$$

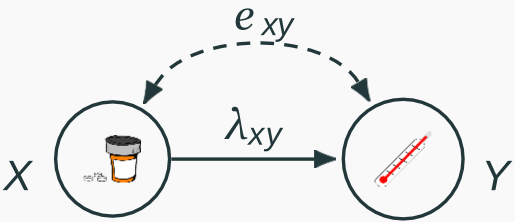
$$Y := \lambda_{xy} X + e_y$$

e_x, e_y correlated

The drug is expensive so mostly rich people are getting it.
But data not gathered...

The Ground-Truth Model

If one is unable to ascertain the assumption of no confounding between X and Y, this is the corresponding graphical model:

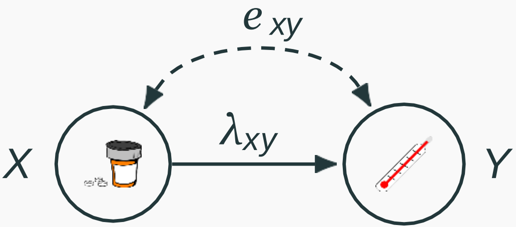


Regression $Y = \beta X + e$ gives *biased answer*

$$\sigma_{xy} = \lambda_{xy}\mathrm{E}[XX] + \mathrm{E}[e_x e_y]$$
$$= \boxed{\lambda_{xy} + e_{xy}}$$

The Ground-Truth Model

If one is unable to ascertain the assumption of no confounding between X and Y, this is the corresponding graphical model:



It is provably impossible to disentangle the effect of the drug from the confounding.

That is, λ_{xy} is **not identifiable**

What does Regression Compute?

$$Y = \beta X + e$$

Here, β is the regression coefficient.

What does β represent?

What does Regression Compute?

Let's do least squares symbolically:

$$\begin{aligned}
\mathbf{E}[(Y-\beta X)^2] &= \mathbf{E}[YY - 2\beta XY + \beta^2 XX] \\
&= \mathbf{E}[YY] - 2\beta\mathbf{E}[XY] + \beta^2\mathbf{E}[XX] \\
&= 1 + \beta^2 - 2\beta\mathbf{E}[XY] \\
&= 1 + \beta^2 - 2\beta\sigma_{xy}
\end{aligned}$$

Minimizing:

$$\begin{aligned}
0 &= \frac{\partial \mathbb{E}[(Y-\beta X)^2]}{\partial\beta} = \frac{\partial}{\partial\beta}\ 1 + \beta^2 - 2\beta\sigma_{xy} \\
&= 2\beta - 2\sigma_{xy} \\
\beta &= \sigma_{xy}
\end{aligned}$$

The regression coefficient is just the covariance between x and y!

Regression Equation vs. SCM: Confusion of the Century

- **Regression Equation:**

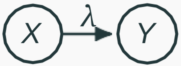
$$Y = \beta X + e \qquad \text{Assuming } e \perp X$$

When solved, $\beta = \sigma_{xy}$. We will call this value r_{yx} (solved value of linear regression of y on x). It makes no causal claims.

- **Structural Equation:**

$$Y = \lambda X + e_y$$

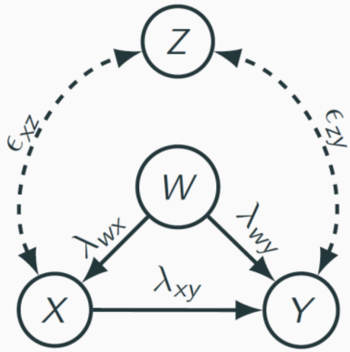
$$\mathbf{E}[Y|do(X)] = \lambda X$$



Makes claims about the interventional distribution which can be tested, and can be falsified.

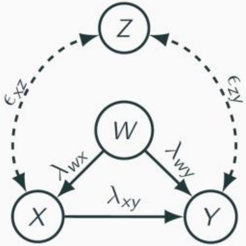
Be Careful With Regression

Remember: alpha, beta are regression Coeffieints and lmbdas aree causal



Be Careful With Regression

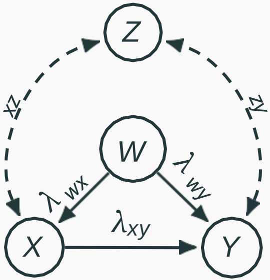
Remember: alpha, beta are regression
Coeffieints and lmbdas aree causal



$$Y = \beta X + e$$

$$\beta = \sigma_{xy} = \lambda_{xy} + \lambda_{wx}\lambda_{wy}$$

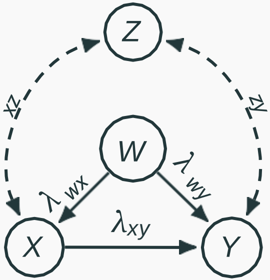
Be Careful With Regression



$$Y = \beta X + \alpha W + \gamma Z + e$$

$$\beta = \lambda_{xy} - \frac{\epsilon_{xz}\epsilon_{zy}}{1-\lambda_{wx}^2-\epsilon_{xz}^2}$$

Be Careful With Regression

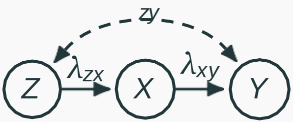


$$Y = \beta X + \alpha W + e$$

$$\beta = \lambda_{xy}$$

How to Use Regression Correctly?

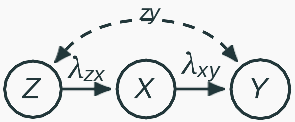
Single-Door Criterion



We want to find λ_{xy}.

$$r_{yx} = \sigma_{xy} = ??$$

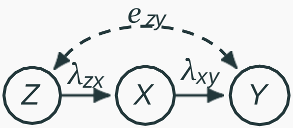
Single-Door Criterion



We want to find λ_{xy}. How can it be isolated?

$$r_{yx} = \sigma_{xy} = \lambda_{xy} + \lambda_{zx}\ e_{zy}$$

Single-Door Criterion: Multiple Regression



What if we find the least squares regression parameters of this model?

$$Y = \alpha X + \beta Z + e$$

$$\alpha = \lambda_{xy}$$
$$\beta = e_{zy}$$

Single-Door Criterion

Theorem Single-Door (Identification of Direct Effects) [8]

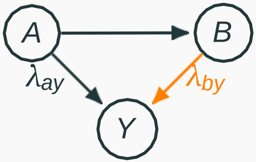
Let G be any path diagram in which λ is the path coefficient associated with the link $X \rightarrow Y$, and let G_λ denote the diagram that results when $X \rightarrow Y$ is removed from G. The coefficient λ is identifiable if there exists a set Z such that

1. Z contains no descendants of Y, and
2. Z D-separates X from Y in G_λ

Moreover, if Z satisfies these conditions, $\lambda = r_{yxz}$

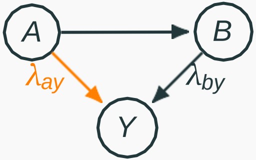
Here, we use the notation r_{yxz} to be the regression coefficient of x when performing regression y on x and z.

Example



$$\lambda_{by} = ?$$

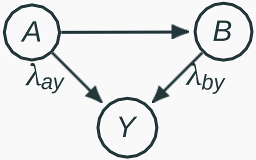
Example



$$\lambda_{by} = r_{yba}$$

$$\lambda_{ay} = ?$$

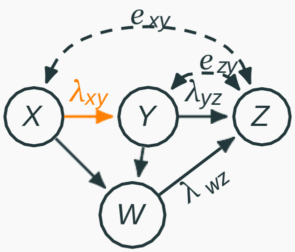
Example



$$\lambda_{by} = r_{yba}$$

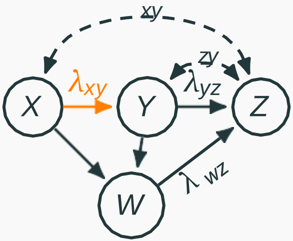
$$\lambda_{ay} = r_{yab}$$

Try It



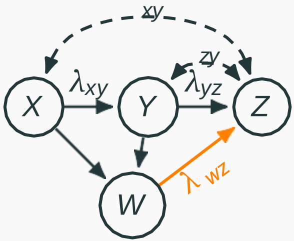
$$\lambda_{xy} = ?$$

Try It



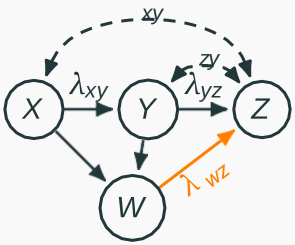
$$\lambda_{xy} = r_{yx}$$

Try It Again



$$\lambda_{wz} = ?$$

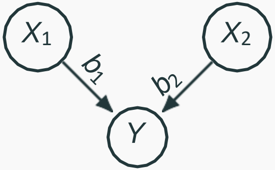
Try It Again



$$\lambda_{wz} = r_{zwyx}$$

Corollary: When are Multiple Parameters Useful?

When can we use multiple regression to solve for multiple coefficients *simultaneously?*



Back-Door Criterion

Theorem Back-Door (Identification of Total Effects) [8]

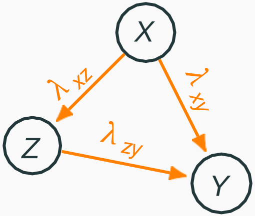
For any two variables X and Y in a causal diagram G, the total effect of X on Y is identifiable if there exists a set of measurements Z such that

1. No member of Z is a descendant of X, and
2. Z d-separates X from Y in the subgraph $G_{\underline{X}}$

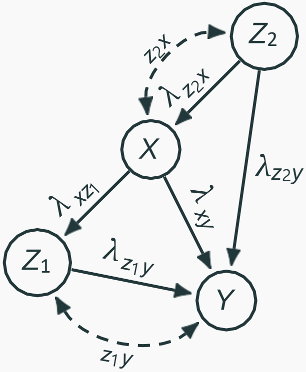
Moreover, if Z satisfies these conditions, the total effect of X on Y is given by r_{yxz}

Remember that $G_{\underline{X}}$ means delete all edges outgoing from X.

Why no Descendants of X?

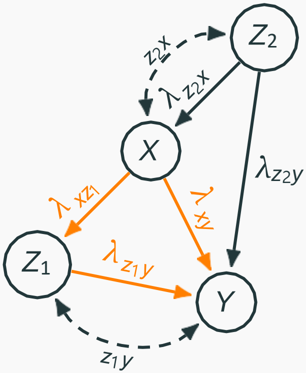


Example



What is the total effect of X on Y?

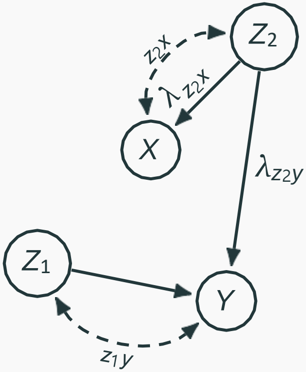
Example



What is the total effect of X on Y? $\lambda_{xz_1}\lambda_{z_1y} + \lambda_{xy}$

Can we find it using the back-door?

Example



What is the total effect of X on Y? $\lambda_{xz_1}\lambda_{z_1y} + \lambda_{xy}$

Can we find it using the back-door? r_{yxz_2}